

Stanford University

Department of Computer Science
Stanford, California, 94305-2140

Terry Winograd
Professor of Computer Science
Telephone: 415 723 2780
Fax: 415 725 7411
Internet: winograd@cs.stanford.edu

Sept. 1, 1994

To the participants in the Stanford World-Wide-Web-Workshop

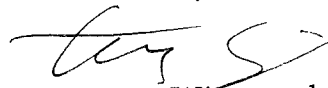
The enclosed information about the workshop is also on the Web (except for the brief overview from the *Communications of the ACM*). If you have access to a Web browser, you should start with the workshop home page URL:

<http://www-pcd.stanford.edu/workshop/workshop.html>

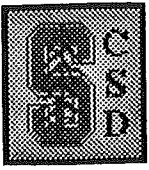
If you have questions about arrangements for the workshop, please contact Carolyn Tajnai <tajnai@cs.stanford.edu>. For questions about the speakers and content, you can ask me.

We're all looking forward to a fun and educational workshop. Thanks for coming.

Sincerely,



Terry Winograd



Schedule for Stanford WWW Workshop

Stanford Computer Forum - September 20-21, 1994

Tuesday, Sept. 20:

8:00 - Continental Breakfast

8:45 - Welcome and introduction

9:00 - Session 1

- Naming and resource location - Larry Masinter, Xerox PARC, masinter@parc.xerox.com
- Search and query protocols - Brewster Kahle, WAIS Inc, brewster@wais.com
- Spiders and autonomous agents - Yoav Shoham, Stanford Nobotics group - shoham@cs.stanford.edu
- Web structure and meta-information - Terry Winograd, Stanford Project on People, Computers, and Design - winograd@cs.stanford.edu

10:30 - Break

11:00 - Discussions

12:00 - lunch

2:00 - Session 2

- Browsers - Dale Dougherty, O'Reilly Associates - dale@ora.com
- A Visual Design Perspective - Ev Shafir, Hewlett-Packard - shafir@hpuid.ptp.hp.com
- Page Layout and portability - Steve Zilles, Adobe - zilles@adobe.com
- Generating virtual pages - Tom Gruber, Stanford Knowledge Systems Laboratory - gruber@hpcs.stanford.edu

3:30 - Break

4:00 - Discussions

5:00 - Informal reception

Wednesday, September 21

8:30 - Continental Breakfast

9:00 - Session 3

- Security - Allan Schiffman, Enterprise Integration Technologies - ams@eit.com
- Using the Internet for commerce - Mike Genesereth, Stanford Logic Group - genesereth@cs.stanford.edu
- Enhancing the WWW with co-presence Ehud Shapiro, Ubique Ltd. - udi@ubique.co.il
- On line communities - Sean White, Interval Research - white@interval.com
- Digital Libraries Hector Garcia-Molina, Stanford Database Group - hector@cs.stanford.edu

10:30 Break

11:00 Discussions

12:00 Planning for bird-of-a-feather sessions in the afternoon

12:30 Lunch break

Wed. p.m.: Open schedule with meeting rooms made

available

In addition to these talks, we will have demonstrations of a number of Web-based systems being developed or used at Stanford. We will have number of workstations available for participants to browse the web and follow up on information resources mentioned in the sessions.



Readings for Stanford WWW Workshop

Stanford Computer Forum - September 20-21, 1994

This information was prepared for participants in the WorldWideWeb Workshop. It is intended as a starting point for further web exploration, and is not comprehensive or definitive (most information on the Web isn't!).

WWW (The World Wide Web)

Brief executive summary

A short introduction to the Web by Tim Berners-Lee, then at CERN, now at MIT, the primary originator of the Web.

An illustrated talk

An online seminar by Berners-Lee on the World-Wide Web (W3). It gives an overview of W3 for those to whom it is new, including a review of the current status of software, and then mentions some plans for the future.

Guide to Cyberspace

Kevin Hughes from EIT wrote this guide (last update, May 1994). It offers a nice introduction to the Web's conceptual structure, to hypermedia, to Mosaic, to HTML and URLs (Uniform Resource Locators), and to the current state of the Web.

World Wide Web FAQ (Frequently Asked Questions)

This page has many pointers to detailed information, and the high-level information that it provides is fairly useful. It includes a short comparison of WWW, Gopher, and WAIS.

World-Wide Web Consortium: A Short Prospectus

The W3O (WWW Organization) is a consortium located at MIT, in collaboration with CERN, which is attempting to play a central role in standardizing and promoting the Web, as the X-Consortium at MIT did with X-Windows.

HTML (The HyperText Markup Language)

A beginners guide to HTML

This is a primer for producing documents in HTML, the markup language used by the World Wide Web.

The official HTML specification

There is much debate over just what "official" means on the Web. The emergence of the WWW consortium may help. For the moment, files produced at CERN (such as this one) are more official than others.

HTML Description

The abstract for the current HTML specification by Dan Connolly (July 1994). It has links to the specification itself (highly detailed), and the abstract itself is informative.

Technical specs for HTML

Detailed Technical specs for those interested in knowing what is really going on.

HTML+ (or HTML 3, as it is now known)

Covers the important elements that will be introduced in planned new versions

SGML (the Standard Generalized Markup Language)

A description of SGML

SGML is the basis for HTML (the Hyper Text Markup Language).

SGML FAQ

This article contains answers to questions that are frequently posted to the news group comp.text.sgml, it is intended for newcomers to the list and SGML beginners.

HTTP (The Hypertext Transfer Protocol)

HTTP technical description

HTTP is a protocol with the lightness and speed necessary for a distributed collaborative hypermedia information system. It is a generic stateless object-oriented protocol, which may be used for many similar tasks such as name servers, and distributed object-oriented systems, by extending the commands, or "methods", used. A feature of HTTP is the negotiation of data representation, allowing systems to be built independently of the development of new advanced representations.

Secure HTTP

Proposal for standards for a version of HTTP that provides security through cryptographic techniques.

Searching

CUSI - Search everything

A metaindex that lets you get to a lot of other indexes of what is on the Web.

The Internet Computer Index

More of the same including a small sample of what kind of stuff is available for ftp'ing on the web broken down by computer type.

Some interesting things to look at

The Virtual Tourist - A nice way of looking at information from a "World-Wide" perspective and shows the use of image maps.

The World-Wide Web Virtual Library: Subject Catalogue - An ambitious attempt by Arthur Secret at Cern, and volunteers, to provide a hierarchical subject catalog of what is on the web.

PLANET EARTH HOME PAGE - Another interesting way of approaching a lot of information, much of it graphical and geographical.

Yahoo page

An ad hoc attempt by some graduate students at Stanford to maintain a hierarchical listing of "cool stuff" on the net.

GNN Home Page (Global Network Navigator)

The first web-based commercial magazine.

The InterNIC InfoGuide Home Page - Information about the Internet and guides to it (including a server that helps you locate people on the net).

The Web at Nexor A number of facilities: list of Archie gateways in the Web. Macintosh Archive in hypertext. ALIWEB: Archie-Like Indexing in the Web. Search engines for RFC's and Internet Drafts. CUSI: a configurable unified search interface. A local copy of the CUI W3 Catalog (July 1st) A mirror of frequently used international files. A section dedicated to The Perl Programming language. A section on World Wide Web Robots, Wanderers and Spiders. References to current satellite images, etc.

Usenet FAQ lists in hypertext Lists of the Frequently Asked Questions postings collected for all of the Internet News groups.

Winograd@cs.stanford.edu

This page is <http://www-pcd.stanford.edu/workshop/readings.html>

World Wide Web Frequently Asked Questions

This document resides on the World Wide Web on: Sunsite (URL is http://sunsite.unc.edu/boutell/faq/www_faq.html).

If you are unfamiliar with the term "URL", read on and learn!

Last update: 9/2/94

Contents

- 1: Recent changes to the FAQ
- 2: Information about this document
- 3: Elementary Questions
 - 3.1: What are WWW, hypertext and hypermedia?
 - 3.2: What is a URL?
 - 3.3: How does WWW compare to gopher and WAIS?
- 4: Accessing the Web (User Questions)
 - 4.1: Introduction: How can I access the web? (Even by email!)
 - 4.2: Browsers Accessible by Telnet
 - 4.3: Obtaining browsers
 - 4.3.1: Microsoft Windows browsers
 - 4.3.2: MSDOS browsers
 - 4.3.3: Macintosh browsers
 - 4.3.4: Amiga browsers
 - 4.3.5: NeXTStep browsers
 - 4.3.6: X/DecWindows (graphical UNIX, VMS) browsers
 - 4.3.7: Text-based Unix and VMS browsers
 - 4.3.8: Batch-mode "browsers"
 - 4.4: How can I access the web through a firewall?
 - 4.5: What is on the web?
 - 4.5.1: How do I find out what's new on the web?
 - 4.5.2: Where is the subject catalog of the web?
 - 4.5.3: How can I search through ALL web sites?
 - 4.6: How can I save an inline image to disk?
 - 4.7: How can I get sound from the PC speaker with WinMosaic?
 - 4.8: I have a Windows PC (or a Macintosh). Why can't I open WAIS URLs?
 - 4.9: I'm running XMosaic. Why can't I get external viewers working?
 - 4.10: Hey, I know, I'll write a WWW-exploring robot! Why not?
 - 4.11: How do I send newsgroup posts in HTML to my web client?
- 5: Providing Information to the Web (Provider Questions)
 - 5.1: Introduction: How can I provide information to the web?
 - 5.2: Obtaining Servers
 - 5.2.1: Unix Servers
 - 5.2.2: Macintosh Servers
 - 5.2.3: Windows and Windows NT Servers
 - 5.2.4: MSDOS Servers

- [5.2.5: VMS Servers](#)
- [5.2.6: Amiga Servers](#)
- [5.3: Producing HTML documents](#)
 - [5.3.1: Writing HTML directly](#)
 - [5.3.2: HTML editors](#)
 - [5.3.3: Converting other formats to HTML](#)
- [5.4: How do I publicize my work?](#)
- [5.5: Can I buy space on an existing server?](#)
- [5.6: Advanced Provider Questions](#)
 - [5.6.1: How do I set up a clickable image map?](#)
 - [5.6.2: How do I make a "link" that doesn't load a new page?](#)
 - [5.6.3: Where can I learn how to create fill-out forms?](#)
 - [5.6.3.1: How can I create hidden fields in forms \(keeping state\)?](#)
 - [5.6.3.2: How can users email me through their browsers?](#)
 - [5.6.4: How do I comment an HTML document?](#)
 - [5.6.5: How can I create decent-looking tables and stop using <PRE>...</PRE>?](#)
 - [5.6.6: What is HTML Level 3 and where can I learn more about it?](#)
 - [5.6.7: How can I make transparent GIFs?](#)
 - [5.6.8: Which format is better for WWW images, JPEG or GIF?](#)
 - [5.6.9: How can I mirror part of another server?](#)
 - [5.6.10: How come mailto: URLs don't work?](#)
 - [5.6.11: How can I restrict and control access to my server?](#)
 - [5.6.12: How can I keep robots off my server?](#)
- [6: What newsgroups discuss the web?](#)
- [7: I want to know more.](#)
- [8: Credits](#)

1: Recent additions and changes to the FAQ

- [9/2/94: Email forms](#)
- [9/2/94: Keeping robots off your server](#)
- [9/2/94: Quadralay commercial-grade Mosaic](#)
- [9/2/94: New location of alternate BBEdit tools](#)
- [9/2/94: Emacs-W3 browser works on the Amiga](#)
- [9/2/94: Enhanced imagemaps section \(URLs for other editors wanted!\)](#)
- [9/2/94: Big Dummy's Guide is now EFF's Guide](#)
- [9/2/94: Fixed location of Postscript HTML tutorial](#)
- [9/2/94: Added Mac program to transparent section](#)
- [9/2/94: Enhanced section on problems with XMosaic external viewers](#)
- [9/2/94: Removed references to obsolete HTML+ draft](#)
- Closed all <A NAME> tags. Should make browsers happier.
- [9/2/94: Updated location of WinMosaic](#)
- [9/2/94: Updated URL of web space leasing document](#)
- [9/2/94: Email access to the web](#)

(Up to Table of Contents)

2: Information about this document

This is an introduction to the World Wide Web project, describing the concepts, software and access methods. It is aimed at people who know a little about navigating the Internet, but want to know more about WWW specifically. If you don't think you are up to this level, try an introductory Internet book such as Ed Krol's "The Whole Internet" or "EFF's Guide to the Internet". The latter is available electronically by anonymous FTP from <ftp:eff.org> in the directory pub/Net_info/EFF_Net_Guide.

This informational document is posted to <news.answers>, <comp.infosystems.www.users>, <comp.infosystems.www.providers>, <comp.infosystems.www.misc>, <comp.infosystems.gopher>, <comp.infosystems.wais> and <alt.hypertext> every four days (please allow a day or two for it to propagate to your site). The latest version is always available on the web as http://sunsite.unc.edu/boutell/faq/www_faq.html. (see the section titled "What is a URL?" to understand what this means.)

The most recently posted version of this document is kept on the <news.answers> archive on <rtfm.mit.edu> in </pub/usenet/news.answers/www/faq>. For information on FTP, send e-mail to mail-server@rtfm.mit.edu with:

send usenet/news.answers/finding-sources
in the body (not subject line) of your message, instead of asking me.

[Thomas Boutell](#) maintains this document. Feedback about it is to be sent via e-mail to boutell@netcom.com.

In all cases, regard this document as out of date. Definitive information should be on the web, and static versions such as this should be considered unreliable at best. The most up-to-date version of the FAQ is the version maintained on the web. Please excuse any formatting inconsistencies in the posted version of this document, as it is automatically generated from the on-line version.

(Up to Table of Contents)

3: Elementary questions

3.1: What are WWW, hypertext and hypermedia?

WWW stands for "World Wide Web". The WWW project, started by [CERN](#) (the European Laboratory for Particle Physics), seeks to build a distributed hypermedia system.

The advantage of hypertext is that in a hypertext document, if you want more information about a particular subject mentioned, you can usually "just click on it" to read further detail. In fact, documents can be and often are linked to other documents by completely different authors -- much like footnoting, but you can get the referenced document instantly!

To access the web, you run a [browser program](#). The browser reads documents, and can fetch documents from other sources. Information providers set up [hypermedia servers](#) which browsers can get documents from.

The browsers can, in addition, access files by [FTP](#), [NNTP](#) (the [Internet news protocol](#)), [gopher](#) and [an ever-increasing range of other methods](#). On top of these, if the server has search capabilities, the browsers will permit [searches of documents and databases](#).

The documents that the browsers display are hypertext documents. Hypertext is text with pointers to other text. The browsers let you deal with the pointers in a transparent way -- select the pointer, and

you are presented with the text that is pointed to.

Hypermedia is a superset of hypertext -- it is any medium with pointers to other media. This means that browsers might not display a text file, but might display images or sound or animations.

[\(Up to Table of Contents\)](#)

3.2: What is a URL?

URL stands for "Uniform Resource Locator". It is a draft standard for specifying an object on the Internet, such as a file or newsgroup.

URLs look like this: (file: and ftp: URLs are synonymous.)

- file://wuarchive.wustl.edu/mirrors/msdos/graphics/gifkit.zip
- ftp://wuarchive.wustl.edu/mirrors
- http://info.cern.ch:80/default.html
- news:alt.hypertext
- telnet://dra.com

The first part of the URL, before the colon, specifies the access method. The part of the URL after the colon is interpreted specific to the access method. In general, two slashes after the colon indicate a machine name (machine:port is also valid).

When you are told to "check out this URL", what to do next depends on your browser; please check the help for your particular browser. For the line-mode browser at CERN, which you will quite possibly use first via telnet, the command to try a URL is "GO URL" (substitute the actual URL of course). In Lynx you just select the "GO" link on the first page you see; in graphical browsers, there's usually an "Open URL" option in the menus.

[\(Up to Table of Contents\)](#)

3.3: How does WWW compare to gopher and WAIS?

While all three of these information presentation systems are client-server based, they differ in terms of their model of data. In gopher, data is either a menu, a document, an index or a telnet connection. In WAIS, everything is an index and everything that is returned from the index is a document. In WWW, everything is a (possibly) hypertext document which may be searchable.

In practice, this means that WWW can represent the gopher (a menu is a list of links, a gopher document is a hypertext document without links, searches are the same, telnet sessions are the same) and WAIS (a WAIS index is a searchable page, returning a document with no links) data models as well as providing extra functionality.

Gopher and World Wide Web usage are now running neck and neck, according to the statistics-keepers of the Internet backbone. (Of course, World Wide Web browsers can also access Gopher servers, which inflates the numbers for the latter.) This is changing as WWW reaches critical mass (usage of the server at CERN doubles every 4 months -- twice the rate of Internet expansion).

[\(Up to Table of Contents\)](#)

4.1: Introduction: how can I access the web?

You have three options: use a browser on your own machine (the best option), use a browser that can be telnetted to (not as good), or access the web by email (the least attractive, but for some it's the only way). It is always best to run a browser on your own machine, unless you absolutely cannot do so; but feel free to telnet to a browser for your first look at the web, or use email if the telnet command does not

work on your system (*try it first!*) The following sections cover telnetting to a browser and obtaining your own browser; if neither of these are possible for you (because you have only an email-and-news connection to the Internet), here is how to access a web page by email:

Send email to listserv@info.cern.ch containing the following single line. (What you put on the subject line doesn't matter; blank is OK. This line should go in the text of the message.) You will receive as a reply a simple page intended to help you learn more about the Web.

send <http://www.earn.net/gnrt/www.html>

(Up to Table of Contents)

4.2: Browsers accessible by telnet

An up-to-date list of these is available on the Web as

<http://info.cern.ch/hypertext/WWW/FAQ/Bootstrap.html> and should be regarded as an authoritative list.

info.cern.ch

No password is required. This is in Switzerland, so continental US users might be better off using a closer browser.

www.cc.ukans.edu

A full screen browser "Lynx" which requires a vt100 terminal. Log in as www. Does not allow users to "go" to arbitrary URLs, so GET YOUR OWN COPY of Lynx and install it on your system if your administrator has not done so already. The best plain-text browser, so move mountains if necessary to get your own copy of Lynx!

www.njit.edu

(or telnet 128.235.163.2) Log in as www. A full-screen browser in New Jersey Institute of Technology. USA.

www.huji.ac.il

A dual-language Hebrew/English database, with links to the rest of the world. The line mode browser, plus extra features. Log in as www. Hebrew University of Jerusalem, Israel.

sun.uakom.cs

Slovakia. Has a slow link, only use from nearby.

info.funet.fi

(or telnet 128.214.6.102). Log in as www. Offers several browsers, including Lynx (goto option is disabled there also).

fserv.kfki.hu

Hungary. Has slow link, use from nearby. Login is as www.

(Up to Table of Contents)

4.3: Obtaining browsers

The preferred method of access of the Web is to run a browser yourself. Browsers are available for many platforms, both in source and executable forms. Here is a list generated from the authoritative list, <http://info.cern.ch/hypertext/WWW/Clients.html>.

(Up to Table of Contents)

4.3.1: Microsoft Windows browsers

NOTE: These browsers require that you have SLIP, PPP or other TCP/IP networking on your PC. SLIP or

PPP can be accomplished over phone lines, but only with the active cooperation of your network provider or educational institution. If you only have normal dialup shell access, your best option at this time is to run Lynx on the Unix (or VMS, or...) system you call, or telnet to a browser if you cannot do so.

Cello

Browser from Cornell LII. Available by anonymous FTP from [ftp.law.cornell.edu](ftp://law.cornell.edu) in the directory /pub/LII/cello.

Mosaic for Windows

From NCSA. Available by anonymous FTP from [ftp.ncsa.uiuc.edu](ftp://ncsa.uiuc.edu) in the directory PC/Windows/Mosaic. (Up to Table of Contents)

4.3.2: MSDOS browsers

NOTE: These browsers require that you have SLIP, PPP or other TCP/IP networking on your PC. SLIP or PPP can be accomplished over phone lines, but only with the active cooperation of your network provider or educational institution. If you only have normal dialup shell access, your best option at this time is to run Lynx on the Unix (or VMS, or...) system you call, or telnet to a browser if you cannot do so.

DosLynx

DosLynx is an excellent text-based browser for use on DOS systems. You must have a level 1 packet driver, or an emulation thereof, or you will only be able to browse local files; essentially, if your PC has an Ethernet connection, or you have SLIP, you should be able to use it. DosLynx can view GIF images, but not when they are inline images (as of this writing). See the README.HTM file at the DosLynx site for details. You can obtain DosLynx by anonymous FTP from [ftp2.cc.ukans.edu](ftp://ftp2.cc.ukans.edu) in the directory pub/WWW/DosLynx; the URL is <ftp://ftp2.cc.ukans.edu/pub/WWW/DosLynx/>.

(Up to Table of Contents)

4.3.3: Macintosh browsers

NOTE: These browsers require that you have SLIP, PPP or other TCP/IP networking on your PC. SLIP or PPP can be accomplished over phone lines, but only with the active cooperation of your network provider or educational institution. If you only have normal dialup shell access, your best option at this time is to run Lynx on the Unix (or VMS, or...) system you call, or telnet to a browser if you cannot do so.

Mosaic for Macintosh

From NCSA. Full featured. Available by anonymous FTP from [ftp.ncsa.uiuc.edu](ftp://ncsa.uiuc.edu) in the directory Mac/Mosaic.

Samba

From CERN. Basic. Available by anonymous FTP from <info.cern.ch> in the directory /ftp/pub/www/bin as the file mac.

MacWeb

From EInet. Has features that Mosaic lacks; lacks some features that Mosaic has. Available by anonymous FTP from [ftp.einet.net](ftp://einet.net) in the directory einet/mac/macweb.

(Up to Table of Contents)

4.3.4: Amiga browsers

AMosaic

Browser for AmigaOS, based on NCSA's Mosaic. Supports older Amigas as well as the newer machines in the latest versions, I am told; available for anonymous ftp from max.physics.sunysb.edu in the directory /pub/amosaic, or from aminet sites in /pub/aminet/comm/net. see the site for details. See the URL <http://insti.physics.sunysb.edu/AMosaic/home.html>.

Emacs-W3

The Emacs-W3 browser works under Gnu Emacs on the Amiga (see section 4.3.7).

(Up to Table of Contents)

4.3.5: NeXTStep browsers

Note: NeXT systems can also run X-based browsers using one of the widely used X server products for the NeXT. The browsers listed here, by contrast, are native NeXTStep applications.

OmniWeb

A World Wide Web browser for NeXTStep. The URL for more information is <http://www.omnigroup.com/>; you can ftp the package from ftp.omnigroup.com in the /pub/software/ directory.

WorldWideWeb, CERN's NeXT Browser-Editor

A browser/editor for NeXTStep. **Currently out of date; editor not operational.** Allows wysiwyg hypertext editing. Requires NeXTStep 3.0. Available for anonymous FTP from info.cern.ch in the directory /pub/www/src.

(Up to Table of Contents)

4.3.6: X/DecWindows (graphical UNIX, VMS) browsers

NCSA Mosaic for X

Unix browser using X11/Motif. Multimedia magic. Full http 1.0 support including PUT-method forms, image maps, etc. Recommended if you can run it. Available by anonymous FTP from <ftp.ncsa.uiuc.edu> in the directory Mosaic.

NCSA Mosaic for VMS

Browser using X11/DecWindows/Motif. For the VMS operating system. Multimedia magic. Full http 1.0 support including PUT-method forms, image maps, etc. Recommended if you can run it. Available by anonymous FTP from <ftp.ncsa.uiuc.edu> in the directory Mosaic.

Quadralay GWHIS Viewer (Commercial Mosaic)

Quadralay offers a commercial-grade (not free!) version of Mosaic for Unix systems, with Windows and Macintosh versions expected in the future. (URL is: <http://www.quadralay.com/products/products.html#gwhis>)

tkWWW Browser/Editor for X11

Unix Browser/Editor for X11. (Beta test version.) Available for anonymous ftp from harbor.ecn.purdue.edu in the directory tkwww[extension] (followed by an extension possibly dependent on the current version). Please ftp to the site and look for the latest version (or use the link above). Supports WYSIWYG HTML editing.

MidasWWW Browser

A Unix/X browser from Tony Johnson. (Beta, works well.)

Viola for X (Beta)

Viola has two versions for Unix/X: one using Motif, one using Xlib (no Motif). Handles HTML Level 3 forms and tables. Has extensions for multiple columnning, collapsible/expandable list,

client-side document include. Available by anonymous FTP from ora.com in /pub/www/viola.

More information available at the URL

<http://xcf.berkeley.edu/ht/projects/viola/README>.

Chimera

Unix/X Browser using Athena (doesn't require Motif). Supports forms, inline images, etc.; closest to Mosaic in feel of the non-Motif X11 browsers. Available for anonymous FTP from <ftp.cs.unlv.edu> in the directory /pub/chimera.

(Up to Table of Contents)

4.3.7: Text-mode Unix and VMS browsers

These are text-based browsers for Unix (and in some cases also VMS) systems. In many cases your system administrator will have already installed one or more of these packages; check before compiling your own copy.

Line Mode Browser

This program gives W3 readership to anyone with a dumb terminal. A general purpose information retrieval tool. Available by anonymous ftp from <info.cern.ch> in the directory /pub/www/src.

The "Lynx" full screen browser

This is a hypertext browser for vt100s using full screen, arrow keys, highlighting, etc. Available by anonymous FTP from <ftp2.cc.ukans.edu>.

Tom Fine's perlWWW

A tty-based browser written in perl. Available by anonymous FTP from <archive.cis.ohio-state.edu> in the directory pub/w3browser as the file w3browser-0.1.shar.

For VMS

Dudu Rashty's full screen client based on VMS's SMG screen management routines. Available by anonymous FTP from <vms.huji.ac.il> in the directory www/www_client.

Emacs w3-mode

W3 browse mode for emacs. Uses multiple fonts when used with Lemacs or Epoch. See the [documentation](#). Available by anonymous FTP from <moose.cs.indiana.edu> in the directory pub/elisp/w3 as the files w3.tar.Z and extras.tar.Z.

(Up to Table of Contents)

4.3.8: Batch-Mode "Browsers"

Batch mode browser

A batch-mode "browser", url_get, which is available through the URL http://wwwhost.cc.utexas.edu/test/zippy/url_get.html. It can be retrieved via anonymous FTP to ftp.cc.utexas.edu, as the file /pub/zippy/url_get.tar.Z. This package is intended for use in cron jobs and other settings in which fetching a page in a command-line fashion is useful.

(Up to Table of Contents)

4.4: How can I access the web through a firewall?

For information on using NCSA Mosaic from behind a firewall, please read the following. In general, browsers can be made useful behind firewalls through the use of a package called "SOCKS"; the source must be modified slightly and rebuilt to accommodate this. Whenever possible, work *with* your

network administrators to solve the problem, not against them.

An excerpt from the NCSA Mosaic FAQ:

NCSA Mosaic requires a direct internet connection to work, but some folks have put together a package that works behind firewalls. This is *completely unsupported* by NCSA, but here is the latest announcement:

November 15, 1993: C&C Software Technology Center (CSTC) of NEC Systems Lab has made available a version of SOCKS, a package for running Internet clients from behind firewalls without breaching security requirements, that includes a suitably modified version of Mosaic for X 2.0. Beware: such a version is not supported by NCSA; we can't help with questions or problems arising from the modifications made by others. But, we encourage you to check it out if it's interesting to you. Questions and problem notifications can be sent to Ying-Da Lee (ylee@syl.dl.nec.com).

[\(Up to Table of Contents\)](#)

4.5: What is on the web?

Currently accessible through the web:

- anything served through gopher
- anything served through WAIS
- anything on an FTP site
- anything on Usenet
- anything accessible through telnet
- anything in hytelnet
- anything in hyper-g
- anything in techinfo
- anything in texinfo
- anything in the form of man pages
- sundry hypertext documents

[\(Up to Table of Contents\)](#)

4.5.1: How do I find out what's new on the web?

The unofficial newspaper of the World Wide Web is [What's New With NCSA Mosaic](http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/whats-new.html) (URL is <http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/whats-new.html>), which carries announcements of new servers on the web and also of new web-related tools. This should be in your hot list if you're not using Mosaic (which can access it directly through the help menu).

[\(Up to Table of Contents\)](#)

4.5.2: Where is the subject catalog of the web?

There are several. There is no mechanism inherent in the web which forces the creation of a single catalog (although there is work underway on automatic mechanisms to catalog web sites). The best-known catalog, and the first, is [The WWW Virtual Library](http://info.cern.ch/hypertext/DataSources/bySubject/Overview.html) (URL is <http://info.cern.ch/hypertext/DataSources/bySubject/Overview.html>), maintained by CERN. The Virtual Library is a good place to find resources on a particular subject, and has separate maintainers for many subject areas.

There is also a newer cataloging system called [ALIWEB](#) that requires very little effort to maintain

and is growing rapidly (URL is <http://web.nexor.co.uk/aliweb/doc/aliweb.html>).

[\(Up to Table of Contents\)](#)

4.5.3: How can I search through ALL web sites?

Several people have written robots which create indexes of web sites -- including sites which have not arranged to be mentioned in the newspapers and catalogs above. (Before writing your own robot, please [read the section on robots.](#))

Here are a few such automatic indexes you can search:

- [WebCrawler](http://www.biotech.washington.edu/WebQuery.html) (URL is <http://www.biotech.washington.edu/WebQuery.html>) builds an impressively complete index; on the other hand, since it indexes the content of documents, it may find many links that aren't exactly what you had in mind.
- [World Wide Web Worm](http://www.cs.colorado.edu/home/mcbryan/WWW.html) (URL is <http://www.cs.colorado.edu/home/mcbryan/WWW.html>) builds its index based on page titles and URL contents only. This is somewhat less inclusive, but pages it finds are more likely to be an exact match with your needs.

You can read about other robots in the [robots section](#).

[\(Up to Table of Contents\)](#)

4.6: How can I save an inline image to disk?

Here are two ways:

1. Turn on "load to local disk" in your browser, if it has such an option; then reload images. You'll be prompted for filenames instead of seeing them on the screen. Be sure to shut it off when you're done with it.
2. Choose "view source" and browse through the HTML source; find the URL for the inline image of interest to you; copy and paste it into the "Open URL" window. This should load it into your image viewer instead, where you can save it and otherwise muck about with it.

[\(Up to Table of Contents\)](#)

4.7: How can I get sound from the PC speaker with WinMosaic?

This piece of wisdom donated by Hunter Monroe:

This section explains how to install sound on a PC which already has a working version of Mosaic for Microsoft Windows. Be warned in advance that the results may be poor.

To get Mosaic to produce sound out of the PC speaker, first, you need a driver for the speaker. You can [get the Microsoft speaker driver](#) from the URL

<ftp://ftp.microsoft.com/Softlib/MSLFILES/SPEAK.EXE> or by doing an Archie search to find it somewhere else. SPEAK.EXE is a self-extracting file. Copy the speak.exe file to a new directory, and then type "SPEAK" at the DOS prompt. Do not put the file SPEAKER.DRV in a separate directory from OEMSETUP.INF.

Now, you need to install the driver. In Windows, from the Program Manager choose successively Main/Control Panel/Drivers/Add/Unlisted or updated drivers/(enter path of SPEAK.EXE)/PC Speaker. At this point some strange sounds come out as the driver is initialized. Change the settings to improve the sound quality on the various sounds: tr da, chimes, etc. Click OK when you are finished and choose the Restart windows option.

Having installed the speaker driver, you will now get sounds whenever you start Windows, make a

mistake, or exit Windows. If you do not want this, from the Main/Control Panel/Sounds menu, make sure there is no X next to "Enable System Sounds."

Now, you need a sound viewer program that Mosaic can call to display sounds. NCSA unfortunately recommend WHAM, which does not work well with a PC speaker. Get the program WPLANY instead. You can find a copy nearby with an Archie search on the string "wplny"; the current version is WPLNY09B.ZIP. For details on archie and other basic issues related to FTP, please read the Usenet newsgroup news.announce.newusers.

Move the zip file to a new directory, and use an unzip program like pkunzip to unzip it, producing the files WPLANY.EXE and WPLANY.DOC. Then edit the MOSAIC.INI file to remove the "REM" before the line "TYPE9=audio/basic". Then, you need lines in the section below that read something like: audio/basic="c:\wplany\wplany.exe %ls" audio/wav="c:\wplany\wplany.exe %ls" where you have filled in the correct path for wplany.exe. The MOSAIC.INI file delivered with Mosaic may have NOTEPAD.EXE on the audio/basic line, but this will not work. Now, restart Mosaic, and you should now be able to produce sounds. To check this, with Mosaic choose File/Local File/\WINDOWS*.WAV and then try to play TADA.WAV. Then, you might try the Mosaic Demo document for some .AU sounds, but you are lucky if your speaker produces something you can understand.

(Up to Table of Contents)

4.8: I have a Windows PC or Macintosh. Why can't I access WAIS URLs?

This answer provided by Michael Grady (m-grady@uiuc.edu):

The version of Mosaic for X has "wais client" code built-in to it. This was relatively easy for the developers to do, because there was already a set of library routines for talking to WAIS available for Unix as "public domain" (freeWAIS). I don't think there is such a library of routines for PC/Windows or Mac, which would make it much more difficult for the Mosaic versions for Windows and the Mac to add "wais client" capability. Therefore, at least for now, neither the Windows or Mac versions of Mosaic support direct query of a WAIS server (i.e. can act as wais clients themselves).

(Up to Table of Contents)

4.9: I'm running XMosaic. Why can't I get external viewers working...

... No matter what no matter what I do to my .mailcap and .mime.types files?

Answer provided by Ronald E. Daniel (rdaniel@acl.lanl.gov):

Mosaic only looks at the .mime.types file if it has no idea what the document's type is. This is actually a very rare situation. Essentially all servers now use the HTTP/1.0 protocol, which means that they tell Mosaic (or other browsers) what the document's MIME Content-type is. The servers use a file very much like Mosaic's .mime.types file to infer the Content-type from the filename's extension.

It is pretty simple to find out if this really is the problem. Use telnet to talk to the server and find out if it is assigning a MIME type to the document in question. Here's an example, looking at the home page for my server. (idaknow: is my shell prompt)

```
idaknow: telnet www.acl.lanl.gov 80 // Connect to the httpd server
Trying 128.165.148.3 ...
Connected to www.acl.lanl.gov.
Escape character is '^]'.
HEAD /Home.html HTTP/1.0 // replace Home.html with your document
// you supply the blank line
HTTP/1.0 200 OK // the rest of this comes from the server Date: Wednesday, 25-May-94
```

```
Server: NCSA/1.1
MIME-version: 1.0
Content-type: text/html           // Here's the MIME Content-type
Last-modified: Monday, 16-May-94 16:21:58 GMT
Content-length: 1727
```

Connection closed by foreign host.

idaknow:

In the example above, /Home.html will get <http://www.acl.lanl.gov/Home.html>.

Normally servers will be configured to supply a Content-type of text/plain if they don't know what else to do. If this is the problem you are having, take a look at the [TypesConfig documentation](#) for [NCSA's httpd](#). You can have the server look at the filename extension, supply the correct Content-type, then use your local .mailcap file to tell Mosaic what viewer to use to look at the document.

Russ Segal adds:

The answer from Ronald Daniel is essentially correct, but it needs a small addendum.

When starting Moasic, you can specify a "fileProxy" which will fetch files for you:

```
"*fileProxy: http://socks/"
```

If you do this, file: URLs are no longer strictly local accesses. So even if the URL is not ftp:, the proxy server must be upgraded as Mr. Daniel suggests.

[\(Up to Table of Contents\)](#)

4.10: Hey, I know, I'll write a WWW-exploring robot! Why not?

Programs that automatically traverse the web can be quite useful, but have the potential to make a serious mess of things. Robots have been written which do a "breadth-first" search of the web, exploring many sites in a gradual fashion instead of aggressively "rooting out" the pages of one site at a time. Some of these robots now produce excellent indexes of information available on the web.

But others have written simple depth-first searches which, at the worst, can bring servers to their knees in minutes by recursively downloading information from CGI script-based pages that contain an infinite number of possible links. (Often robots can't realize this!) Imagine what happens when a robot decides to "index" the CONTENTS of several hundred mpeg movies. Shudder.

The moral: a robot that does what you want may already exist; if it doesn't, please study the document [World Wide Web Robots, Wanderers and Spiders](#) (URL is: <http://web.nexor.co.uk/mak/doc/robots/robots.html>) and learn about the emerging standards for exclusion of robots from areas in which they are not wanted. You can also read about existing robots there.

[\(Up to Table of Contents\)](#)

4.11: How do I send newsgroup posts in HTML to my web client?

How to do this depends greatly on your system; if you have a Mac or Windows system, the answer is completely different. But, as food for thought, here is a simple shell script I use on my Unix account to send posts from rn and related newsreaders to [Lynx](#). Put this text in the file "readwebpost" and use the "chmod" command to make it executable, then put it somewhere in your path (such as your personal bin directory):

```
#!/bin/sh
```

```

echo \<PRE\> > .article.html
cat >> .article.html
echo \</PRE\> >> .article.html
lynx .article.html < /dev/tty
rm .article.html

```

Then add the following line to your .rnmac file (create it if you don't already have one):

```
W      |readwebpost %C
```

Now, when you press "W" while reading a post in rn, a message will be sent to Lynx, and the links enclosed in it will be live.

Larry W. Virden provides the following version which invokes Mosaic instead, and is also capable of communicating with an already-running copy of Mosaic instead of launching another. (You can use the same rn macro as above, invoking "goto-xm" instead of "readwebpost".) Read the comments for details on the assumptions made by the script.

```

#!/bin/sh
# goto-xm, by Joseph T. Buck
# Modified heavily by Larry W. Virden
# Script for use with newsreaders such as trn.  Piping the article
# through this command causes xmosaic to pop up, pointing to the
# article.  If an existing xmosaic (version 1.1 or later) exists,
# the USR1 method will be used to cause it to point to the correct
# article, otherwise a new one will be started.

# assumptions: ps command works as is on SunOS 4.1.x, may need changes
# on other platforms.

URL=`/bin/grep '^Message-ID:' | /bin/sed -e 's/.*</news:/' -e 's/>.*//'`
if [ "X$URL" = "X" ]; then
    echo "USAGE: $0 [goto] [once] < USENET_msg" >&2
    exit 1
fi

pid=`ps -xc | egrep '[Mm]osaic' | awk 'NR == 1 {print $1}'`
p=`which Mosaic`
gfile=/tmp/Mosaic.$pid

$P "$URL" &

if [ "$#" -gt 0 ]; then
    if [ "$1" = "goto" -o "$1" = "same" ]; then
        shift
        echo "goto" > $gfile
    else
        echo "newwin" > $gfile
    fi
else
    echo "newwin" > $gfile
fi

/bin/awk 'END { printf "'"$URL"'" }' </dev/null >> $gfile

```

```
trap "echo signal encountered" 30
```

```
kill -USR1 $pid
```

```
exit 0
```

[\(Up to Table of Contents\)](#)

5.1: Introduction: How can I provide information to the web?

Information providers run programs that the browsers can obtain hypertext from. These programs can either be WWW servers that understand the HyperText Transfer Protocol HTTP (best if you are creating your information database from scratch), "gateway" programs that convert an existing information format to hypertext, or a non-HTTP server that WWW browsers can access -- anonymous FTP or gopher, for example.

To learn more about World Wide Web servers, you can consult a [www server primer by Nathan Torkington](#), available at the URL <http://www.vuw.ac.nz/who/Nathan.Torkington/ideas/www-servers.html>.

If you only want to provide information to local users, placing your information in local files is also an option. This means, however, that there can be no off-machine access.

[\(Up to Table of Contents\)](#)

5.2: Obtaining Servers

Servers are available for Unix, Macintosh, MS Windows, and VMS systems. If you know of a server for another operating system, please contact me.

See <http://info.cern.ch/hypertext/WWW/Daemon/Overview.html> for more information on writing servers and gateways in general.

[\(Up to Table of Contents\)](#)

5.2.1: Unix Servers

NCSA httpd

NCSA has released a server, known as the NCSA httpd; it is available at the URL ftp://ftp.ncsa.uiuc.edu/Web/ncsa_httpd.

CERN httpd

CERN's server is available for anonymous FTP from info.cern.ch (URL is <http://info.cern.ch/hypertext/WWW/Daemon/Status.html>) and many other places. Use your local copy of archie to search for "www" in order to find a nearby site.

GN Gopher/HTTP server

The GN server is unique in that it can serve both WWW and Gopher clients (in their native modes). This is a good server for those migrating from Gopher to WWW, although it does not have the server-side-script capabilities of the NCSA and CERN servers. See the URL <http://hopf.math.nwu.edu/>.

Perl server

There is also a server written in the Perl scripting language, called [Plexus](#), for which documentation is available at the URL <http://bsd.com/server/doc/plexus.html>.

[\(Up to Table of Contents\)](#)

5.2.2: Macintosh Servers

There is a server for the Macintosh, [MacHTTP](#), available at the URL

There is a server for the Macintosh, MacHTTP, available at the URL
http://www.uth.tmc.edu/mac_info/machttp_info.html.

(Up to Table of Contents)

5.2.3: MS Windows and Windows NT Servers

HTTPS (Windows NT)

HTTPS is a server for Windows NT systems, both Intel and Alpha -- based. It is available via anonymous FTP from [emwac.ed.ac.uk](ftp://emwac.ed.ac.uk/pub/https) in the directory `pub/https` (URL is <ftp://emwac.ed.ac.uk/pub/https>). (Be sure to download the version appropriate to your processor.) You can read a detailed announcement at the FTP site, or by using the URL <ftp://emwac.ed.ac.uk/pub/https/https.txt>.

NCSA httpd for Windows

The NCSA httpd for Windows has most of the features of the Unix version, including scripts (which generate pages on the fly based on user input). It is available by anonymous FTP from [ftp.ncsa.uiuc.edu](ftp://ftp.ncsa.uiuc.edu) in the `Web/ncsa_httpd/contrib` directory as the file `wh1p11a6.zip`, or at the URL ftp://ftp.ncsa.uiuc.edu/Web/ncsa_httpd/contrib/wh1p11a6.zip.

SerWeb

A simple, effective server for Windows written by Gustavo Estrella. Available by anonymous FTP from [winftp.cica.indiana.edu](ftp://winftp.cica.indiana.edu) (or one of its mirror sites, such as [nic.switch.ch](ftp://nic.switch.ch)), as the file `serweb03.zip`, in the directory `/pub/pc/win3/winsock`.

There is also a Windows NT version of SerWeb, available by anonymous FTP from [emwac.ed.ac.uk](ftp://emwac.ed.ac.uk/pub/serweb/serweb_i.zip) as `/pub/serweb/serweb_i.zip`.

WEB4HAM

Another Windows-based server, available by anonymous FTP from [informatik.uni-hamburg.de](ftp://informatik.uni-hamburg.de) as `/pub/net/winsock/web4ham.zip`.

(Up to Table of Contents)

5.2.4: MSDOS Servers

KA9Q NOS (`nos11c.exe`) is a internet server package for DOS that includes HTTP and Gopher servers. It can be obtained via anonymous FTP from one of the following sites:

[inorganic5.chem.ufl.edu](ftp://inorganic5.chem.ufl.edu)

[biochemistry.cwru.edu](ftp://biochemistry.cwru.edu)

(Up to Table of Contents)

5.2.5: VMS Servers

CERN HTTP for VMS

A port of the CERN server to VMS. Available at the URL
[http://delonline.cern.ch/disk\\$user/duns/doc/vms/distribution.html](http://delonline.cern.ch/disk$user/duns/doc/vms/distribution.html).

Region 6 Threaded HTTP Server

A native VMS server which uses DECthreads(tm). This is a potentially major performance advantage because VMS has a high overhead for each process, which is a problem for the frequently-forking NCSA and CERN servers that began life under Unix. A multithreaded server avoids this overhead. Available at the URL
<http://kcgl1.eng.ohio-state.edu/www/doc/serverinfo.html>.

[\(Up to Table of Contents\)](#)

5.2.6: Amiga Servers

NCSA's Unix server has been ported to the Amiga, and is bundled with the AMosaic browser. See the URL <http://insti.physics.sunysb.edu/AMosaic/home.html> for details.

[\(Up to Table of Contents\)](#)

5.3: Producing HTML documents

HTML is the simple markup system used to create hypertext documents. There are three ways to produce HTML documents: writing them yourself, which is not a very difficult skill to acquire, using an HTML editor, which assists in doing the above, and converting documents in other formats to HTML. The following three sections cover these possibilities in sequence.

[\(Up to Table of Contents\)](#)

5.3.1: Writing HTML documents yourself

You can write an HTML document with any text editor. Try the "source" button of your browser (or "save as" HTML) to look at the HTML for a page you find particularly interesting. The odds are that it will be a great deal simpler than you would expect. If you're used to marking up text in any way (even red-pencilling it), HTML should be rather intuitive.

[A beginner's guide to HTML](#) is available at the URL <http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html>. You can also find a plain text version (at the URL <ftp://ftp.ncsa.uiuc.edu/ncsapubs/WWW/HTMLPrimer.txt>) and a compressed Postscript version (at the URL <ftp://ftp.ncsa.uiuc.edu/ncsapubs/WWW/HTMLPrimer.ps.Z>). (Since the latter two are FTP URLs, you can fetch them by hand using FTP if you do not yet have a web browser.)

There is also a [good set of HTML documentation](#) available at the URL <http://www.ucc.ie/info/net/html/doc.html>.

There is also an [HTML primer by Nathan Torkington](#) at the URL <http://www.vuw.ac.nz/who/Nathan.Torkington/ideas/www-html.html>.

[\(Up to Table of Contents\)](#)

5.3.2: HTML editors

Of course, most folks would still prefer to use a friendlier, graphical editor. Some editors are WYSIWYG (What You See Is What You Get), or close to it; others simply assist you in writing HTML by plugging in the desired markup tags for you from a menu.

Fans of the EMACS editor can use EMACS and [html-helper-mode](#), an EMACS "mode" for HTML editing (URL is <http://www.reed.edu/~nelson/tools/>).

There is also another Emacs HTML mode, [html-mode.el](#) (URL is <ftp://ftp.ncsa.uiuc.edu/Web/elisp/html-mode.el>).

For Microsoft Windows users, there is an editor called HTML Assistant with features to assist in the creation of HTML documents. It can be had by anonymous FTP from <ftp.cs.dal.ca> in the directory [/htmlasst/](#). Read the README.1ST file in this directory for information on which files to download.

A WYSIWYG editor for the Web, *SoftQuad HoTMetaL*, is available for downloading at NCSA and

other Mosaic server sites. Many mirror sites exist; if you can't get through to one, try another, don't give up! That's what mirror sites are for. (Also be sure to use the copy closest to you geographically if possible.)

Known mirrors:

- <ftp://ftp.ncsa.uiuc.edu/Mosaic/contrib/SoftQuad/sqhotmetal-1.0.tar.gz>
- <ftp://ftp.ifi.uio.no/pub/SGML/HoTMetaL>
- <ftp://sgml1.ex.ac.uk/SoftQuad>
- <ftp://doc.ic.ac.uk/pub/packages/WWW/ncsa/contrib/SoftQuad>
- <ftp://askhp.ask.uni-karlsruhe.de/pub/infosystems/mosaic/contrib/SoftQuad>
- <ftp://ftp.cs.concordia.ca/pub/www>

You need a Sun SPARC or Microsoft Windows system and 6MB of disk (6MB of RAM minimum for MS Windows). Because it is context-sensitive, HoTMetaL guides users in creating new HTML documents and in cleaning up old ones. A Publish command changes appropriate SRC and HREF attributes from local paths to http locations. For more information, FTP the README file from the same directory, or send email to hotmetal@sq.com. A HoTMetaL Pro commercially supported version is available for purchase from SoftQuad and its resellers.

An editor for all X users: TkWWW (listed above under X browsers) supports WYSIWYG HTML editing; and since it's a browser, you can try out links immediately after creating them.

Also for X users, there is a package called htmltext which supports WYSIWYG HTML editing. More information is available at the URL <http://web.cs.city.ac.uk/homes/njw/htmltext/htmltext.html>.

For Macintosh users, there is evidently a near-WYSIWYG package called HTML Editor (URL is http://dragon.acadiau.ca:1667/~giles/HTML_Editor).

Also for Macintosh users, the BBEdit HTML extensions allow the BBEdit and BBEdit Lite text editors for the Macintosh to conveniently edit HTML documents. (URL is <http://www.uji.es/bbedit-html-extensions.html>.) You can also obtain the extensions package by anonymous ftp from [sumex-aim.stanford.edu](ftp://sumex-aim.stanford.edu) as `info-mac/bbedit-html-ext-b3.hqx`.

There is an alternative BBEdit extension package available as well (URL is <http://www.york.ac.uk/~ld11/BBEditTools.html>). it is available by FTP from [ftp.york.ac.uk](ftp://ftp.york.ac.uk) in the directory `/pub/users/ld11/BBEdit_HTML_Tools.sea.hqx`.

NCSA's List of Filters and Editors, for which the URL is <http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/faq-software.html#editors>, mentions several editors, including two for MS Windows. In some cases, the "editor" amounts to a set of macros for an existing word processor, which can provide a near-WYSIWYG environment.

Note that this URL contains uppercase and lowercase letters; certain operating systems won't allow mixed case on the command line, or will only allow it if it is quoted (VMS), so if you are launching Lynx or another client and specifying a URL at the command line, try quoting the URL in double-quotes ("URL").

Another option, if you have an SGML editor, is to use it with the HTML DTD.

(Up to Table of Contents)

5.3.3: Converting other formats to HTML

There is a collection of filters for converting your existing documents (in TeX and other non-HTML formats) into HTML automatically, including filters that can allow more or less WYSIWYG editing using various word processors:

Rich Brandwein and Mike Sendall's List at CERN. The URL is <http://info.cern.ch/hypertext/WWW/Tools/Filters.html>.

(Note that this URL contains uppercase and lowercase letters; certain operating systems such as VMS require you to quote mixed-case URLs when launching a browser from the command line. This is NOT a bug in the browser.)

There is also a Word for Windows template for writing HTML documents, available at the URL http://www.gatech.edu/word_html/release.htm.

(Up to Table of Contents)

5.4: How do I publicize my work?

There are several things you can do to publicize your new HTML server or other offering:

- Submit it to the NCSA What's New Page at the URL <http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/whats-new.html> (see the page for details on how to submit your listing!).
- Post it to the newsgroup comp.infosystems.announce. Please read the group first to get a feel for the contents. You should not post to comp.infosystems.www.users, misc, providers, etc., but if you feel compelled to do so, please choose misc as announcements are of interest to both providers and users (and those who wear both hats).
- Submit it to the maintainers of various catalogs, such as the WWW Virtual Library (at the URL <http://info.cern.ch/hypertext/DataSources/bySubject/Overview.html>) and the ALIWEB index (at the URL <http://web.nexor.co.uk/aliweb/doc/aliweb.html>).

(Up to Table of Contents)

5.5: Can I buy space on an existing server?

Yes, you can. A list of sites offering WWW space for lease is available (at the URL <http://union.ncsa.uiuc.edu/www/leasing.html>).

(Up to Table of Contents)

5.6.1: How do I set up a clickable image map?

There are really two issues here: how to indicate in HTML that you want an image to be clickable, and how to configure your server to do something with the clicks returned by Mosaic, Chimera, and other clients capable of delivering them.

You can read about image maps and the NCSA server at the URL <http://hoohoo.ncsa.uiuc.edu/docs/setup/admin/Imagemap.html>.

Using imagemaps requires that you create a map file; you can do this by hand or with a WYSIWYG tool. I wrote Mapedit (URL is: <http://sunsite.unc.edu/boutell/mapedit/mapedit.html>), which is such a tool for Microsoft Windows and the X Window System. Other tools are available. (URLs, anyone?)

Important Note: Creating imagemaps requires a cooperative server administrator and a real web server. Don't waste time making maps before making sure you have the necessary tools to deliver them.

(Up to Table of Contents)

5.6.2: How do I make a "link" that doesn't load a new page?

Such links are useful when a form is intended to perform some action on the server machine without sending new information to the client, or when a user has clicked in an undefined area in an image map; these are just two possibilities.

Rob McCool of NCSA provided the following wisdom on the subject:

Yechezkal-Shimon Gutfreund (sg04@gte.com) wrote:

: Ok, here is another bizzare request from me:

: I am currently running scripts which I "DO NOT" want to return
: any visible result. That is, not text/plain, not text/HTML, not
: image/gif. The entire results are the side effects of the
: script and nothing should be returned to the viewer.

: It would be nice to have an internally supported null viewer
: so that I could do this, more "cleanly" (ok, ok, I hear your groans).

HTTP now supports a response code of 204, which is no operation. Some browsers such as Mosaic/X 2.* support it. To use it, make your script a nph script and output an HTTP/1.0 204 header. Something like:
HTTP/1.0 204 No response Server: Myscript/NCSA httpd 1.1

(You can learn more about nph scripts from the NCSA server documentation at the URL <http://hoohoo.ncsa.uiuc.edu/docs/>.) Essentially they are scripts that handle their own HTTP response codes.

(Up to Table of Contents)

5.6.3: Where can I learn how to create fill-out forms?

You can read about the Common Gateway Interface (at the URL <http://hoohoo.ncsa.uiuc.edu:80/cgi/>). In addition to documenting the standard interface for which scripts can now be written for both NCSA and CERN-derived servers, these pages also cover HTML forms and how to handle the results on the server side. See the section on email forms for a simple solution to the most commonly desired form.

(Up to Table of Contents)

5.6.3.1: How can I create hidden fields in forms (keeping state)?

Use INPUT TYPE=hidden. An example:

```
<INPUT TYPE=hidden NAME=state VALUE="hidden info to be returned with form">
```

By now, most if not all browsers can handle the hidden type. Note that "hidden" doesn't mean "secret"; the user can always click on "view source".

(Up to Table of Contents)

5.6.3.2: How can users send me email through their browsers?

If you have access to the server's configuration files, or if your server administrator permits users to create their own CGI scripts, you can arrange it. I've written a simple email forms package (URL is: <http://siva.cshl.org/email/index.html>), which does it in ANSI C. There is also a package floating around in Perl (URL, anyone?).

(Up to Table of Contents)

5.6.4: How do I comment an HTML document?

Use the <!-- tag at the beginning of EACH line commented out; close this for EACH line with the -->

tag. Note that comments do not nest, and the sequence "--" may not appear inside a comment except as part of the closing --> tag.

You should **not** try to use this to "comment out" HTML that would otherwise be shown to the user, since some browsers (notably Mosaic) will still pay attention to tags inside the comment and close it prematurely.

Thanks to Joe English for clearing up this issue.

(Up to Table of Contents)

5.6.5: How can I create decent-looking tables and stop using <PRE>... </PRE>?

Tables are a standard feature in HTML Level 3, a new version of HTML. Unfortunately, they are at present implemented only by the Viola and Emacs-W3 browsers, to my knowledge.

However, there is a way to use HTML Level 3 tables now and convert them automatically to HTML, allowing you to design proper tables and install those pages directly when table support arrives in the majority of clients. You can do this using the html+tables package, by Brooks Cutter (bcutter@paradyne.com), which is available for anonymous ftp from sunsite.unc.edu in the directory pub/packages/infosystems/WWW/tools/html+tables.shar. This package requires the shell language Perl, which is primarily used on Unix systems but is also available for other systems (such as MSDOS machines). html+tables accepts HTML Level 3 and outputs html using the <PRE>...</PRE> construct to represent tables, allowing you to write HTML Level 3 now, knowing that it will look better when clients are ready for it.

(Up to Table of Contents)

5.6.6: What is HTML Level 3 and where can I learn more about it?

HTML Level 3, also known as HTML+, is an enhanced version of HTML designed to address some of the limitations of HTML. HTML Level 3 supports true tables, right-justified text, centered text, line breaks that do not double space, and many other desired features.

However, most clients support only a handful of HTML Level 3 features (such as forms in Mosaic) at this time.

You can access information about new developments in HTML at the CERN server (at the URL <http://info.cern.ch/hypertext/WWW/Markup/Markup.html>).

(HTML Level 1 is the original version. HTML Level 2 is essentially the same, but with the addition of forms.)

(Up to Table of Contents)

5.6.7: How can I make transparent GIFs?

Transparent GIFs are useful because they appear to blend in smoothly with the user's display, even if the user has set a background color that differs from that the developer expected.

There is a document explaining transparent GIFs available at the URL http://melmac.corp.harris.com/transparent_images.html. You can fetch the program giftrans by anonymous ftp from ftp.rz.uni-karlsruhe.de at the path /pub/net/www/tools/giftrans.c.

There is also a utility for the Macintosh, Transparency (URL is: <http://www.med.cornell.edu/~giles/projects.html#transparency>).

(Up to Table of Contents)

5.6.8: How come mailto: URLs don't work?

The `mailto:` URL is an innovation found in Lynx and a few other browsers. It is not yet found in Mosaic, the most popular browser. Hopefully it will be present in future versions. In the meantime, you can set up forms which send mail to you; there is documentation on this at the URL <http://siva.cshl.org/email/index.html>.

[\(Up to Table of Contents\)](#)

5.6.9: How can I restrict and control access to my server?

All major servers have features that allow you to limit access to particular sites, and many clients have authentication features that allow you to identify specific users. There is a [tutorial on security and user authentication](#) with the NCSA server and Mosaic available, written by Marc Andreessen (URL is <http://wintermute.ncsa.uiuc.edu:8080/auth-tutorial/tutorial.html>). See your server documentation for further information.

[\(Up to Table of Contents\)](#)

5.6.10: Which format is better for WWW image purposes, JPEG or GIF?

JPEG does a better job with realistic images such as scanned photographs. Most browsers cannot handle inline JPEGs, however, so you must link to them as external images (using a regular `<A HREF...>` instead of ``).

GIF does a better job with crisp, sharp images, such as those typically used to construct buttons, graphs and the like. All browsers that can display graphics at all can display GIFs inline.

[\(Up to Table of Contents\)](#)

5.6.11: How can I mirror part of another server?

Scripts are available to do this, but at this time they are not very friendly to the server you are attempting to mirror; their behavior resembles that of the more poorly written [WWW robots](#). If you are trying to improve access times to a distant server, you will likely find the "proxy" capabilities of [CERN's WWW server](#) to be a more effective and general solution to your problem.

[\(Up to Table of Contents\)](#)

5.6.12: How can I keep robots off my server?

Programs that automatically traverse the web can be quite useful, but have the potential to make a serious mess of things. Every so often someone will write a "depth-first" searching robot that brings servers to their knees. See the section on [writing robots \(4.10\)](#) for details.

Fortunately, most robots on the web follow a simple protocol by which you can keep them off your server if you wish, or keep them out of portions of your server which are robot traps (ie, they contain an infinite number of possible links). Read the document [World Wide Web Robots, Wanderers and Spiders](#) (URL is: <http://web.nexor.co.uk/mak/doc/robots/robots.html>) and learn about the emerging standards for exclusion of robots from areas in which they are not wanted. You can also read about existing robots there, including useful cataloging robots you probably do *not* want to keep off your server.

[\(Up to Table of Contents\)](#)

6: What newsgroups discuss the Web?

You can find discussion of World Wide Web topics in three newsgroups, and one newsgroup which will soon be removed:

<comp.infosystems.www.users>

A forum for the discussion of WWW client software and its use in contacting various Internet information sources. New user questions, client setup questions, client bug reports, resource-discovery questions on how to locate information on the web that can't be found by the means detailed in the FAQ and comparison between various client packages are among the acceptable topics for this group. Please specify what browser and what system type (Windows, Mac, Unix, etc.) your post is about if you are asking questions about a specific program.

comp.infosystems.www.providers

A forum for the discussion of WWW server software and the use of said software to present information to users. General server design, setup questions, server bug reports, security issues, HTML page design and other concerns of information providers are among the likely topics for this group.

comp.infosystems.www.misc

A forum for general discussion of WWW (World Wide Web)- related topics that are NOT covered by the other newsgroups in the hierarchy. This will likely include discussions of the Web's future, politicking regarding changes in the structure and protocols of the web that affect both clients and servers, et cetera.

comp.infosystems.www (DEFUNCT)

The old catch-all newsgroup, which may still exist on your system but will be removed on September 7th, according to David Tale, moderator of news.announce.newgroups.

(Up to Table of Contents)

7: I want to know more

To find out more, use the web. This FAQ hopefully provides enough information for you to locate and install a browser on your system. If you have system specific questions regarding FTP, networking and the like, please consult newsgroups relevant to your particular hardware and operating system!

Once you're up and running, you may wish to consult the World Wide Web Primer by Nathan Torkington. It is available at the URL
<http://www.vuw.ac.nz/who/Nathan.Torkington/ideas/www-primer.html>.

Later you may return to this FAQ for answers to some of the more advanced questions. I encourage you to check out the changes listed early in the document each time the FAQ appears.

(Up to Table of Contents)

8: Credits

- Thomas Boutell boutell@netcom.com
- Nathan Torkington Nathan.Torkington@vuw.ac.nz
- Marc Andreessen marca@ncsa.uiuc.edu
- Tony Johnson

(Up to Table of Contents)

The World-Wide Web

The World-Wide Web (W3) was developed to be a pool of human knowledge, which would allow collaborators in remote sites to share their ideas and all aspects of a common project. Physicists and engineers at CERN, the European Particle Physics Laboratory in Geneva, Switzerland, collaborate with many other institutes to build the software and hardware for high-energy physics research. The idea of the Web was prompted by positive experience of a small "home-brew" personal hypertext system used for keeping track of personal information on a distributed project. The Web was designed so that if it was used independently for two projects, and later relationships were found between the projects, then no major or centralized changes would have to be made, but the information could smoothly reshape to represent the new state of knowledge. This property of scaling has allowed the Web to expand rapidly from its origins at CERN across the Internet irrespective of boundaries of nations or disciplines.

If you haven't yet experienced the Web, the best way to find out about it is to try it. An Appendix to this article gives some recipes for getting hold of W3 clients. Given one of these, you will quickly find out all you need to know, and much more. For hard copy to read on the plane, or if you don't have Internet access from your desktop machine, refer to our paper in *Electronic Networking* (see "Glossary and Further Reading") for an overview of the project, material which we will not repeat but will summarize here.

A W3 "client" program runs on your computer. When it starts, it displays an object, normally a document with text and possibly images. Some of the phrases and images are highlighted: in blue, or boxed, or perhaps numbered, depending on what sort of a display you have and how your preferences have been set. Clicking the mouse on the highlighted area

("anchor") causes the client program to retrieve another object from some other computer, a "server." The retrieved object is normally also in a hypertext format, so the process of navigation continues (see Figure 1).

When viewing some documents, the reader can request a search, by typing in plain text (or complex commands) to send to the server, rather than following a link. In either case, the client sends a request off to the server, often a completely different machine in some other part of the world, and within (typically) a second, the related information, in either hypertext, plain text or multimedia format, is presented. This is done repeatedly, and by a sequence of selections and searches one can find anything that is "out there." Some important things to note are:

- Whatever type of server, the user interface is the same, so users do not need to understand the differences between the many protocols in common use. Before W3, access to networked information typically involved knowledge of many different access "recipes" for different systems, and a different command language for each. The model of hypertext with text input has proved sufficiently powerful to express all the user interfaces, while being sufficiently simple to require no training for a computer user.
- Links can point to anything that can be displayed, including search result lists. (When a query is applied to an object, the resulting object has an address, defined to be the address of the queried object concatenated with the text of the query. As the result object has an address, one can make links to it. Following the link later leads to a reevaluation of the query.)
- While menus and directories are available, the extra option of hypertext provides a more powerful communications tool. In simple cases, the server program can generate a hypertext view representing (for exam-

ple) the directory structure of an existing file store. This allows existing data to be put "on the Web" without further human effort.

- There is a very extendable system for introducing new formats for multimedia data.
- There are many W3 client programs. As hypertext information is transmitted on the network in logical (mark-up) form, each client can interpret this in a way natural for the given platform, making optimal use of fonts, colors, and other human interface resources available on that platform.

What Does W3 Define?

W3 has come to stand for a number of things, which should be distinguished. These include

- The idea of a boundless information world in which all items have a reference by which they can be retrieved;
- The address system (URI) which the project implemented to make this world possible, despite many different protocols;
- A network protocol (HTTP) used by native W3 servers giving performance and features not otherwise available;
- A markup language (HTML) which every W3 client is required to understand, and is used for the transmission of basic things such as text, menus and simple on-line help information across the net;
- The body of data available on the Internet using all or some of the preceding listed items.

The client-server architecture of the Web is illustrated in Figure 2.

Universal Resource Identifiers

Universal Resource Identifiers¹ (URIs) are the strings used as ad-

¹The Internet Engineering Task Force (IETF) is currently defining a similar and derived syntax known as a Uniform Resource Locator (URL). As this work is not complete, and there is no guarantee that URIs will have the same syntax or properties as URLs, we use the term URI here to avoid confusion.



dresses of objects (e.g., menus, documents, images) on the Web. For example, the URI of the main page for the WWW project happens to be

<http://info.cern.ch/hypertext/WWW/TheProject.html>

URIs are "Universal" in that they encode members of the universal set of network addresses. For a new network protocol that has some concept of object, one can form an address for any object as the set of protocol parameters necessary to access the object. If these parameters are encoded into a concise string, with a prefix to identify the protocol and encoding, one has a new URI scheme. There are URIs for Internet news articles and newsgroups (the NNTP protocol), and for FTP archives, for telnet destinations, email addresses, and so on. The same can be done for names of objects in a given name space.

The prefix "http" in the preceding example indicates the address space, and defines the interpretation of the rest of the string. The HTTP protocol is to be used, so the string contains the address of the server to be contacted, and a substring to be passed to the server. Different protocols use different syntaxes, but there is a small amount of common syntax. For example, the common URI syntax reserves the "/" as a way of representing a hierarchical space, and "?" as a separator between the address of an object and a query operation applied to it. As these forms recur in several information systems, to allow expression of them in the common syntax allows the features to be retained in the common model, where appropriate. Hierarchical forms are useful for hypertext, where one "work" may be split up into many interlinked documents. Relative names exploit the hierarchical structure and allow links to be made within the work independent of the higher parts of the URI such as the server name.

URI syntax allows objects to be addressed not only using HTTP, but also using the other common networked information protocols in use today (FTP, NNTP, Gopher, and WAIS), and will allow extension when new protocols are developed.

URIs are central to the W3 archi-

ture. The fact that it is easy to address an object anywhere on the Internet is essential for the system to scale, and for the information space to be independent of the network and server topology.

Hypertext Transfer Protocol

Perhaps misnamed, rather than being a protocol for transferring hypertext, HTTP is a protocol for transferring information with the efficiency necessary for making hypertext jumps. The data transferred may be plain text, hypertext, images, or anything else.

When a user browses the Web, objects are retrieved in rapid succession from often widely dispersed servers. For small documents, the limitations to the response time stem mainly from the number of round trip delays across the network necessary before the rendition of the object can be started. HTTP is therefore a simple request/response protocol.

HTTP does not only transfer HTML documents. Although HTML comprehension is required of W3 clients, HTTP is used for retrieving documents in an unbounded and extensible set of formats. To achieve this, the client sends a (weighted) list of the formats it can handle, and the server replies with data in any of those formats that it can produce. This allows proprietary formats to be used between consenting programs in private, without the need for standardization of those formats. This is important both for high-end users who share data in sophisticated forms, and also as a hook for formats that have yet to be invented. The same negotiation system is used for natural language (English, French, for example) where available, as well as for compression forms.

HTTP is an Internet protocol. It is similar in its readable, text-based style to the File Transfer (FTP) and Network News (NNTP) Protocols that have been used to transfer files and news on the Internet for many years. Unlike these protocols, however, HTTP, is stateless. (That is, it runs over a TCP connection that is held only for the duration of one operation.) The stateless model is efficient when a link from one object may lead equally well to an object stored on the

same server, or to another distant server. The purpose of a reference such as a URI is that it should always refer to the "same" (in some sense) object. This also makes a stateless protocol appropriate, as it returns results based on the URI but irrelevant of any previous operations performed by the client.

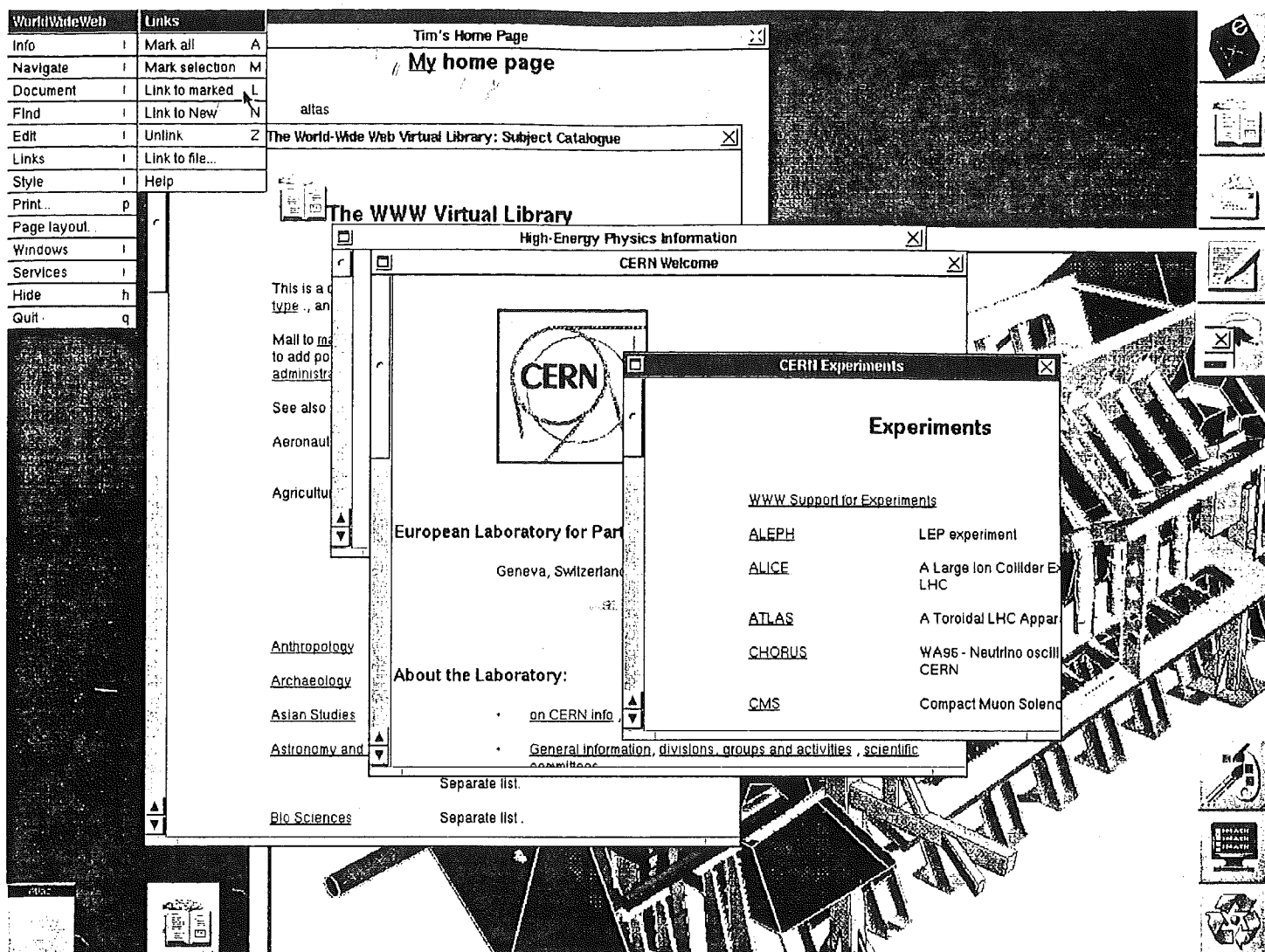
The HTTP request from the client starts with an operation code (known as the method, in conformance with object-oriented terminology) and the URI of the object. The "GET" method used by all browsers is defined to be idempotent in that it should preserve the state of the Web (apart from billing for the information transfer, and statistics). A "PUT" method is defined for front-end update, and a "POST" method for the attachment of a new document to the Web, or submission of a filled-in form or other object to some processor. Use of PUT and POST is currently limited, partly due to scarcity of hypertext editors. The extension to other methods is a subject of study.

When objects are transferred over the network, information about them ("metainformation") is transferred in HTTP headers. The set of headers is an extension of the Multipurpose Internet Mail Extensions (MIME) set. This design decision was taken to open the door to integration of hypermedia mail, news, and information access. Unlike in email, transfer in binary, and transfer in nonstandard but mutually agreed document formats is possible. This allows, for example, servers to indicate links from, and titles of, documents (such as bit-map images) whose data format does not otherwise include such information.

The convention that unrecognized HTTP headers and parameters are ignored has made it easy to try new ideas on working production servers. This has allowed the protocol definition to evolve in a controlled way by the incorporation of tested ideas.

Hypertext Markup Language (HTML)

Despite the ability of HTTP to negotiate formats, W3 needed a common basic language of interchange for hypertext. HTML is that language, and much of the fabric of the Web is constructed out of it. It was designed



to be sufficiently simple so as to be easily produced by both people and programs, but also to adhere to the SGML standard in that a valid HTML document, if attached to SGML declarations including the HTML "DTD," may be parsed by an SGML parser. HTML is a markup language that does not have to be used with HTTP. It can be used in hypertext email (it is proposed as a format for MIME), news, and anywhere basic hypertext is needed. It includes simple structure elements, such as several levels of headings, bulleted lists, menus and compact lists, all of which are useful when presenting choices, and in on-line documents.

Under development is a much enriched version of HTML known as HTML+. This includes features for more sophisticated on-line documentation, form templates for the entry of data by users, tables and mathematical formulae. Currently many brows-

Figure 1. Using the World-Wide Web. Shown here is the authors' prototype World-Wide Web application for NextStep machines. The application initially displays the user's "home" page (top) of personal notes and links (top). Clicking on underlined text takes the reader to new documents. In this case, the user visited the Virtual Library, and, in the high energy physics department, found a link to CERN. Linked to CERN was the "Atlas" collaboration's web including an engineering drawing (background). To save having to follow the same path again, the link menu (shown) allows a new link to be made, for example from text typed into the home page, directly to the Atlas information.

ers support a subset of the HTML+ features in addition to the core HTML set.

HTML is defined to be a language of communication, which actually flows over the network. There is no requirement that files are stored in HTML. Servers may store files in other formats, or in variations on HTML that include extra information of local interest only, and then generate HTML on the fly with each request.

W3 and Other Systems

Two other systems, WAIS (from Thinking Machines Corporation and now WAIS, Inc.) and Gopher (from the University of Minnesota), share W3's client-server architecture and a certain amount of its functionality. Table 1 indicates some of the differences.

The WAIS protocol is influenced largely by the z39.50 protocol designed for networking library catalogs. It allows a text-based search,

Table 1. A comparison of three popular network information projects.

Registered server figures taken April 27, 1993 and April 15, 1994. WAIS: from Thinking Machines Corporation directory, number of distinct hosts. Gopher: from "All the Gophers in the world" register at the University of Minnesota. W3: from Geographical registry at CERN. In all cases many more servers exist which are not directly registered, so these are a very rough guide with no indication of quantity or quality of information at each host.

	WAIS	Gopher	World-Wide Web
Original target application	Text-based information retrieval	Campus-wide information (CWIS)	Collaborative work
Typical objects Text Menus, Graphics Hypertext	YES NO NO	YES YES NO	YES YES YES
Search functions Text search Relevance feedback Reference to other servers	YES YES NO	YES NO YES	YES NO YES
Registered servers April 1993 April 1994	113 137	455 1410	62 829



and retrieval following a search. Indexes to be searched are found by searching in a master index. This two-stage search has been demonstrated to be sufficiently powerful to cover the current world of WAIS data. There are no navigational tools to allow the reader to be shown the available resources, however, or guided through the data: the reader is "parachuted in" to a hopefully relevant spot in the information world, but left without context.

Gopher provides a free text search mechanism, but principally uses menus. A menu is a list of titles, from which the user may pick one. While gopher space is in fact a web containing many loops, the menu system gives the user the impression of a tree. The Veronica server provides a master index for gopher space.

The W3 data model is similar to the gopher model, except that menus are generalized to hypertext documents. In both cases, simple file servers generate the menus or hypertext directly from the file structure of a server. The W3 hypertext model gives the program more power to communicate the options available to the reader, as it can include headings and various forms of list structure, for example, within the hypertext.

All three systems allow for the provision of graphics, sound and video, although because the WAIS system only has access by text search, text has to be associated with graphics files to allow them to be found.

W3 clients provide access to servers of all types, as a single simple interface to the whole Web is considered very important. Unknown to the user, several protocols are in use behind the scenes. A common code library "libwww" put into the public domain by CERN has promoted this uniformity. Whereas one would not wish to see greater proliferation of protocols, the existence of more than one protocol probably allows for the most rapid progress during this phase in the development of the field. It also allows a certain limited confidence that, if an architecture can encompass older systems and allow transition to current systems, it will, by induction, be able to provide a transition to newer and better ideas as they are invented.

Recent W3 Developments

This article, like others in this issue, was derived from material written in April 1993 for the INET'93 conference. Growth of the Web since that time has been so great that this sec-

tion has been completely rewritten. There are now 829 (May: 1,248) rather than 62 registered HTTP servers, and many more client programs available as then.

The initial prototype W3 client was a "wysiwyg" hypertext browser/editor using NeXTStep. We developed a line mode browser, and were encouraging the developments of a good browser for X workstations. One year ago, NCSA's Mosaic W3 browser was in wide use on X workstations. Its easy installation and use was a major reason for the spread of the Web. Today there are many browsers available for workstations, Macintosh and IBM/PC compatible machines, and for users with character-based terminals. Of the latter category, "Lynx" from the University of Kansas provides full-screen access to the Web for users with character terminals or emulators running on personal computers. Since new software is appearing frequently, readers are advised to check the lists on the Web for those most suited to their needs.

The availability of browsers and the availability of quality information have provoked each other. One available indicator of growth has been Merit Inc.'s count of the traffic of various different protocols across the

NSF T3 backbone in the U.S. (see Figure 3).

An indicator of the uptake rate of clients is the load on the *info.cern.ch* W3 server at CERN, which provides information about the Web itself, which more than doubled every 4 months over the three years between April 1991 and April 1994.

Information providers have also blossomed. Some of these provide simple overviews of what is available at particular institutes or in particular fields. Others use the power of the W3 model to provide a virtual world of great richness. Examples of servers that use hypertext in interesting ways are the RAL-Durham Particle Database, and the Legal Information Institute's hypertexts of several great tomes of American law. Franz Hoesel's hypertext version of the Vatican's Renaissance Culture exhibit at the Library of Congress set an example that was followed by many collections of art, history and other fields. The Palo Alto town hall runs a server with everything from building regulations to restaurants. As an example of the increasing use of the Web for commerce, a user-friendly virtual clothing store prompts for one's size, and points to a virtual store containing only those clothes that are the right size and also in stock.

The Future

The W3 initiative occupies the meeting point of many fields of technology. Users put pressure and effort into bringing about the adoption of W3 in new areas. Apart from being a place of communication and learning, and a new market place, the Web is a show ground for new developments in information technology. Some of the developments that we look forward to in the next few years include

- The implementation of a name service that will allow documents to be referenced by name, independent of their location;
- Hypertext editors allowing nonexpert users to make hypertext links to organize published information. This will bring the goal of computer-supported collaboration closer, with front-end update, and annotation;
- More sophisticated document type definitions providing for the needs of

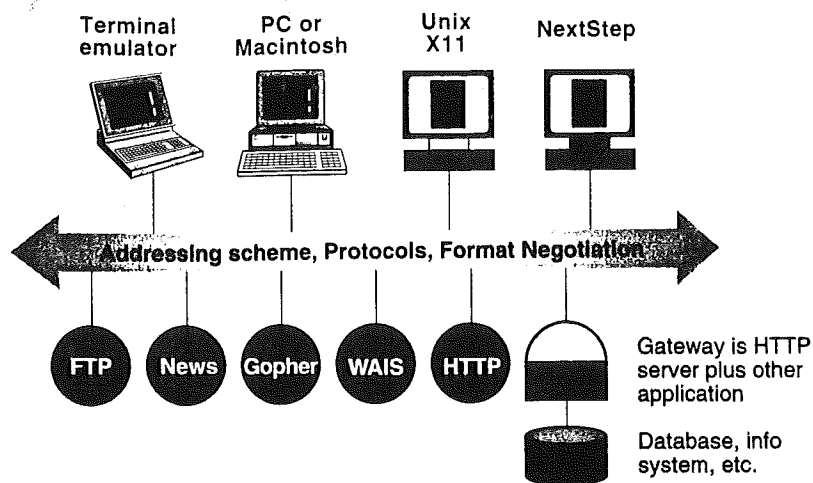


Figure 2. The World-Wide Web client-server architecture. For published information to be universally available, W3 relies on a common addressing syntax, a set of common protocols, and negotiation of data formats.

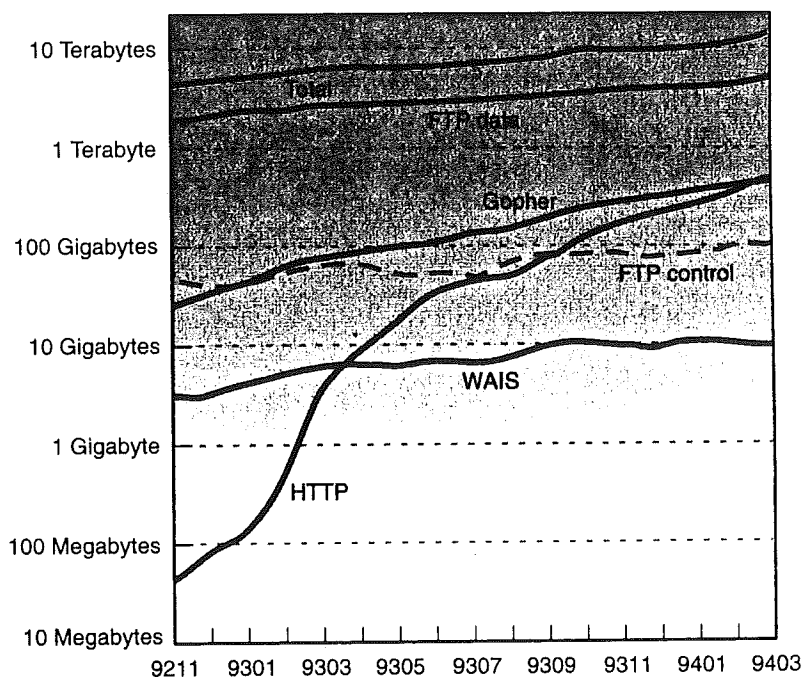
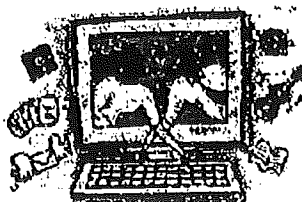


Figure 3. Traffic in bytes per month across the NSF T3 backbone in the U.S. File Transfer Protocol (FTP) was traditionally used to access archives of software. FTP uses separate connections for control and data flow. WAIS arose as an interface to text retrieval systems, Gopher protocol with menu-style interfaces, and W3's HTTP with hypertext and multimedia. W3 clients handle many protocols to access all these worlds of data as a seamless continuum, but new W3 servers use HTTP by preference. Each vertical division represents a tenfold increase in traffic. The horizontal divisions are months. Data: Merit < ftp://ftp.merit.edu/statistics/nsfnet >





Glossary and Further Reading

FTP: File Transfer Protocol. Postel, J. and Reynolds, J. File Transfer Protocol. Internet RFC 959, October 1985. <<ftp://ds.internic.net/rfc/rfc959.txt>>
Gopher: The Internet Gopher. Anklesaria, F. et. al. The Internet Gopher Protocol. Internet RFC 1436, March 1993. <<ftp://ds.internic.net/rfc/rfc1436.txt>>
HTML: Hypertext Markup Language. Berners-Lee, T., and Connolly, D. Hypertext Markup Language Protocol. <<ftp://info.cern.ch/pub/www/doc/html-spec.ps,.txt>>
HTTP: Hypertext Transfer Protocol. Berners-Lee, T. Hypertext Transfer Protocol. <<ftp://info.cern.ch/pub/www/doc/http-spec.ps,.txt>>
MIME: Multipurpose Internet Mail Extensions. Borenstein, N., and Freed, N. MIME (Multipurpose Internet Mail Extensions): Mechanisms for Specifying and Describing the Format of Internet Message Bodies. Internet RFC 1341, June 1992.
NNTP: Network News Transfer Protocol. Kantor, B. and Lapsley, P. A proposed standard for the transmission of news. Internet RFC 977, 1986.
URI: Universal Resource Identifier. Berners-Lee, T. Universal Resource Identifiers for the World-Wide Web. Submitted as an Internet RFC as yet unnumbered. See <<http://info.cern.ch/hypertext/WWW/Addressing/Addressing.html>> for pointers to information on this area.
WAIS: Wide Area Information Servers. See Addyman, T. WAIS: Strengths, Weaknesses and Opportunities. In *Proceedings of Information Networking 93* (London, May 1993), Meckler, London.
W3: Berners-Lee, T.J., Cailliau, R., Groff, J-F, Pollermann, B. World-Wide Web: The Information universe. *Electronic Networking: Research, Applications and Policy*, (Spring 1992), 52-58. See also documents in <<ftp://info.cern.ch/pub/www/doc>> and information referenced by <<http://info.cern.ch/hypertext/WWW/TheProject.html>>

commercial publishers of on-line material;

- The development of a common format for hypertext links from two- and three-dimensional images giving more exciting interface possibilities;
- Integration with concurrent editors and other real-time features such as teleconferencing and virtual reality;
- Easy-to-use servers for low-end machines to ease publication of information by small groups and individuals;
- Evolution of objects from being principally human-readable documents to contain more machine-oriented semantic information, allowing more sophisticated processing;
- Conventions on the Internet for charging and commercial use to allow direct access to for-profit services.

Conclusion

It is intended that after reading this article you will have an idea of what W3 is, where it fits in with other systems in the field, and where it is going. There is much more to be said, especially about providing information, but this is described

on the Web itself. Also in the "Web about the Web" are lists of contributed research and development work and ideas, and pointers to work in progress, so that those interested can work together.

The Web does not yet meet its design goal as being a pool of knowledge that is as easy to update as to read. That level of immediacy of knowledge sharing waits for easy-to-use hypertext editors to be generally available on most platforms. Most information has in fact passed through publishers or system managers of one sort or another. However, the incredible diversity of information available gives great credit to the creativity and ingenuity of information providers, and points to a very exciting future.

Appendix. Getting Started

If you have a vt100 terminal, you can try out a full-screen interface by telnet to ukanaix.cc.ukans.edu and logging in as www. With any terminal, you can telnet to info.cern.ch for the simplest interface. These browsers are also available in source and in some cases binary form. Details of status and coordinates of about 20 differ-

ent browsers are available on the Web—just follow a link to World-Wide Web, and select "software available."

The kernel W3 code (a common code library, and basic server and clients) from CERN is in the public domain. (All protocols and specifications are public domain.) It is available by anonymous FTP from info.cern.ch

NCSA's "Mosaic" browser for W3 is available for X, Mac or PC/Windows by anonymous FTP from ftp.ncsa.uiuc.edu, currently without charge for academic users.

About the Authors:

TIM BERNERS-LEE originated the World-Wide Web in 1990 to enable the sharing of knowledge by complex distributed teams. At CERN he coordinates W3 development by collaborating with institutes around the world. Current research interests include text processing, graphics, communications software, and system design. email: timbl@info.cern.ch

ROBERT CAILLIAU coordinates the use of W3 by CERN experiments and other physics institutes. He is a long-time user of HyperCard, and has been working on W3 since 1991, contributing many ideas, and some software for the Macintosh. email: cailliau@www.cern.ch

ARI LUOTONEN is a member of CERN's technical student program in conjunction with his studies at Tampere University of Technology, Finland. Current research interests include developing CERN's "httpd" HTTP server for Unix and VMS systems. email: luotonen@www.cern.ch

HENRIK FRYSTYK NIELSEN, of Aalborg University, Denmark, is also a CERN technical student. He is working on the kernel code, with research interests in enhanced networking protocols. email: frystyk@info.cern.ch

ARTHUR SECRET wrote the first gateway giving W3 access to a relational database in 1992, while studying Computer Science at Ecole Internationale des Sciences du Traitement de l'Information in Paris, France, as a CERN technical student. Among other tasks in the CERN W3 team, he currently organizes the cataloging of new W3 material in the "virtual library." email: secret@info.cern.ch

Authors' Present Address: CERN, 1211 Geneva 23, Switzerland.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© ACM 0002-0782/94/0800 \$3.50